International Journal of Engineering & Technology, 7 (3.20) (2018) 322-328

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Master Data Domains Prioritization and Privacy Classification in Government Agencies: An empirical study of Malaysia Local Government

[1]Faizura Haneem, [2]Nazri Kama, [3]Azri Azmi, [4]Azizul Azizan and [5]Suriani Mohd Sam

[12345]Advanced Informatics School, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur
**Corresponding Author Email: mdnazri@utm.my*

Abstract

Master data is the most valuable information in each government agency such as customers, suppliers, products, accounts, and the relationships between them. In government agencies, master data are scattered across various agencies, managed in a silo environment in multiple applications and database systems. This silo data management issue led to data quality issues such as data duplication and also may cause disaster to the agencies due to the complexity and higher cost and resource requirement. To resolve the issue, the Master Data Management (MDM) implementation at central level could minimize the data duplication problem by establishing a single source of truth by consolidating and integrating multiple master data from multiple agencies into central repository. The central repository can be referred by other applications across multiple agencies instead of creating new entities in their local database systems. However, in the initial stage of the MDM implementation process, master data across government agencies must be assessed and prioritized to ensure the success of the implementation. Hence, this study aims to clarify the domains prioritization and the privacy classification of master data in government agencies by using a qualitative and quantitative data analysis approach. It involves participative case studies from seven (7) Malaysia's local government agencies. The study identifies 36 sets of master data which generally prioritized into three domains which are; 1) services and products, 2) customers, and 3) service providers. From these master datasets, 20 datasets (56%) are classified as open data. The result of this study is in contrast with most of the current literatures that stated the MDM typically prioritize on the customer data domain as compared to other domains. This study also indicates that the government agency has a high potential to share these open master data to the centralized MDM platform with worry-free of the privacy issues.

Keywords: *data management, master data, master data management (MDM), government agencies*

1. Introduction

Background

Master data describe the most relevant business entities or domains in the agency such as customer, products or suppliers (Otto, Hüner, & Österle, 2012). In most government agencies, the problem in managing master data is that the master data are scattered across various agencies, which comprise of multiple applications and database systems. Master data which have similar information (i.e. individual profile, agency's corporate information) are duplicates across agencies since they have been stored and managed in silos by each agency. Whenever there is a need of data exchange, it is hard-coded in the integration layer (Bonnet, 2013). The silo management of data in each agency raised the issue of data quality (i.e. inconsistent and inaccurate) master data throughout different government agencies which may cause to the disaster to the agencies due to the complexity and higher cost and resource requirement (Bonnet, 2013; Services, 2008).

Master Data Management (MDM) is one of the initiatives by the government to minimize the silo management issues and data quality issues (Loshin, 2009). With MDM, the master data from different agencies that are essential and valuable across government agencies will be recognized and consolidated into a central system. These central master data can be used as a 'single reference of truth' by many applications and systems across government agencies despite creating a new set of master data within their agency (Spruit & Pietzka, 2014). Furthermore, the centralized master data that are classified as open data can serve as the foundation of open data initiatives which provide citizens with better visibility into open government initiatives.

However, according Silvola, Jaaskelainen, Kropsu-Vehkapera, & Haapasalo (2011), defining and prioritizing master data domains across government agencies in initial stage of MDM implementation process is tough. Although customer, product, locations, finance and employee domains have been among the most commonly targeted master data domains where MDM initiatives are focused (Allen & Delton Cervo, 2015), yet the definition and prioritization of the master data domain in government agencies might be different due to the uniqueness of the public sector environment. As such, one of

the challenges in implementing MDM in the government sector is some of the government agencies are reluctant to share and exchange their data due to fear of being exploited by competitors or opponents (Clifton et al., 2004). Hence, this paper aims to clarify the definition of master domains in government agencies and the prioritization in government agencies, and also to define the rate of privacy classification (i.e. open or confidential) of master data in government agencies. In this study, the focus group interviews were conducted to identify the master data domains and the prioritization, and the privacy classification from seven (7) local government agencies as the case studies.

Master Data Management (MDM)

Master Data Management term is being described initially in 2003 by Gartner in one of the research articles in analyzing the SAP solution which is called SAP Master Data Management (SAP MDM). The intention of the solution is to assist the organisation in consolidating and harmonizing their master data, which described as products, locations, customers and suppliers (Gartner, 2003). The MDM solution is designed to be a central repository of master data across multiple systems in order to encourage business process integration and inter-enterprise visibility (Smith & McKeen, 2008).

Master Data

Master data describe the most relevant business entities or domains in the agency such as customer, products or suppliers (Otto et al., 2012). They describe the key characteristics of objects in the real world. A single master data entity is rarely being changed, for instance the properties of some kind of material. Instances of master data classes are relatively unchanged, compared with transactional data. Master data are the reference for transactional data where there would not be a single transactional data without master data.

Additionally, Dreibelbis defines master data is some of the most valuable information about the business domains such as customers, suppliers, products, accounts, and the relationships between them. Furthermore, master data defined as an enterprise data where each of these domains represents information that is needed across different business processes, across agency's units, and between operational systems and decision support systems (Dreibelbis, 2008).

Open Data

The W3C eGov Interest Group defines open data as publishing data in its raw format, machine-readable for its full reuse in applications developed by others (W3C(e-Gov), 2009). The concept of understandable by machines is the publication of data with its embedded semantics so that machines can interpret them in the semantic context in which they were defined.

On the other hand, Open Government Data (OGD) is the publication of government data in open raw formats and ways that make it accessible and readily available to all and allow reuse, such as the creation of data mashup applications (W3C(e-Gov), 2009). Data mashups merge data from two or more different applications or data sources and producing comparative views of the combined information.

Government

The Government sector is usually composed of agencies that are owned and operated by the government, which includes federal, provincial, state, or municipal governments. All other definitions expressed in dictionaries and the literatures indicate public sector organisation to be the agency or body through which a political unit exercises authority and performs functions and usually classified according to the distribution of power within it ("Merriam-Webster Dictionary," 2011). In contrast, the private sector is usually composed of organisations that are privately owned and not part of the government. These usually include corporations (both profit and non-profit) and partnerships.

The government aims to serve people by fulfilling their daily needs, but private sector enterprises are established with profit motive. This different goal may lead to the unique definition and prioritization of master data domains in the government sector as compared to other sectors.

2. Methods

This study was applying qualitative and quantitative data analysis approach to explore and clarify (Linstone & Turoff, 2002; Ranjit Kumar, 2011) the master data domains in government by answering the following main questions:

i. What are the master data domains in government agencies?
ii. From i., what is the priority level (based on the impact to the organisation) of each identified master data?
iii. From i., what is the privacy classification (i.e open or confidential) of each identified master data?

The study was started with literature review and it is followed by conducting focus group interviews from seven (7) local government agencies as case studies. Figure 1 shows the process in conducting this study.

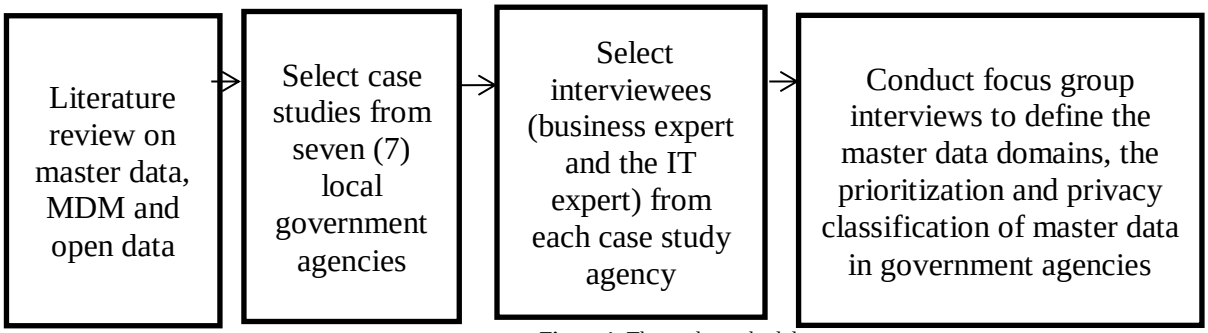

Figure 1: The study methodology

The first process in this study was reviewing the existing literature regarding master data, MDM and open data. From the literature review, the protocol of the focus group interviews is created which includes the component such as the heading, instruction, questionnaires, thank you statement and logs (Creswell, 2014).

Subsequently, seven (7) local government agencies were selected as case studies for focus group interview session. From each selected agency, the business expert and the IT expert were identified as interviewees for the focus group interview. These interviewees were selected depending on their experience and deep understanding with regards to the master data management in their agency. Table 1 indicates the selected government agencies characteristic and the interviewees' roles in each agency. Two (3) focus groups were performed in which each of the focus group consist of three (3) to four (4) participants and led through one (1) moderator. After the focus group interview sessions were conducted and the interviews were recorded, the result of the interviews was analyzed by the researcher.

Table 1: Type of agencies and interviewees' roles from each agency

Agency	Type of agency	Interviewees' roles in agency
Agency 1	State Authority Agency	- Head of Corporate & Communication Department - Head of IT Unit
Agency 2	State Revenue Collecting Agency	- Executive of Collection Management Department - Head of IT Department
Agency 3	Charity Agency	- Executive of Administration Department - Executive of IT Department
Agency 4	Religious Authority Agency	- Executive of Management Services Department - IT Executive
Agency 5	State Syari'a Judiciary Agency	- Deputy Registrar - IT Executive
Agency 6	Local University	- Registrar - Head of IT Department
Agency 7	Municipal Council	- Executive of Procurement Department - Head of IT Department

3. Results

This section describes the result and analysis that consists of two (3) sub-sections which are; 1) master data domains in local government agencies, 2) prioritization of master data domains in local government agencies, and 3) privacy classification of master data domains in local government agencies.

The interviews found out that the government agencies have the common challenges in defining master data in their organisations. Even though the interview questions focused on the master data, the interviewees some time confused between master data, reference data and transactional data. This is due to there is no standard of data definition across multiple agencies in the government sector to categorize whether the data is master data, reference data or transactional data. In addition, data classification system did not exist in some of the selected organisation due to the cost implications. However, to support the interview session, the evaluation of master data in each agency were also based on the data models provided by the organisation which consist of conceptual, logical, and physical models that organize and document the business concepts, data entities, and data elements and the relationships between them.

Master data domains in local government agencies

There are 36 master data entities were recorded by the researcher from the focus group interviews. Overall, the study found that master data entities in local government were generally categorized into three (3) main domains which are 1) customers, 2) services and products, and 3) service providers. Table 2 describes the master data entities and the domains that were recorded from the focus group interviews.

Table 2. Master Data entities and the domains in local government agencies

Agency	Id	Master Data entities	Master Data Domains
Agency 1	MD1	Corporate Information	Services and products
	MD2	Citizens' Profile in the state	Customers
	MD3	Procedures	Services and products
Agency 2	MD4	Corporate Information	Service providers
	MD5	Payment Forms	Services and products
	MD6	Payers' Profile	Customers
	MD7	Receivers' Profile	Customers
	MD8	Procedures	Services and products
Agency 3	MD9	Corporate Information	Service providers
	MD10	Donation Forms	Services and products
	MD11	Donators' Profile	Customers
	MD12	Donation Receivers' Profile	Customers
	MD13	Asset Developers	Customers
	MD14	Procedures	Services and products
Agency 4	MD15	Religious Law & Rules	Services and products

	MD16	Religious Law experts	Service providers
Agency 5	MD17	Judges' Profile	Service providers
	MD18	Lawyers' Profile	Service providers
	MD19	Plaintiffs' Profile	Customers
	MD20	Defendants' Profile	Customers
	MD21	Case Forms	Services and products
	MD22	Procedures	Services and products
Agency 6	MD23	Corporate Information	Service providers
	MD24	Application Forms	Services and products
	MD25	Programs Offered	Services and products
	MD26	Lecturers' Profile	Customers
	MD27	Students' Profile	Customers
	MD28	Procedures	Services and products
	MD29	Students' Feedback/Complaints	Customers
Agency 7	MD30	Customers' Profile	Customers
	MD31	Contractors' Profile	Customers
	MD32	Business License	Services
	MD33	Business Forms	Services
	MD34	Businesses' Profile	Customers
	MD35	Procedures	Services
	MD36	Customers' Feedback/Complaints	Customers

Customers refer to the individual or company that deals with the agency in using the agency services (e.g. Citizens) or provide the third party services to the agency such as contractors or vendors. The Customers domain includes the customers' profile such as demographic, geographic, and psychographic characteristics, services, usage patterns and customer's feedback. Customer profiles are very critical to the agency and normally are used across business unit in the agency. According to the interview result in Table 2, master data that falls under this category are - Agency 1 (citizen profile in the state), Agency 2 (payer profile, receiver profile), Agency 3 (donators profile, receivers profile, asset developers) Agency 5 (plaintiff profile, defendant profile), Agency 6 (lecturers profile, students profile, students' feedback/complaints) and Agency 7 (customers' profile, contractors' profile, businesses' profile, customers' feedback/complaints).

Services and products in public sector refer to the production, delivery and allocation of goods and services to its citizens. The Services and Products domain includes the type of services provided by the agency, service delivery channel (e.g. counter, online form, mobile app) and the procedures involved (e.g. forms, checklists). From the interview result in Table 2, master data that fall under this category are - Agency 1 (procedures), Agency 2 (payment forms, procedures), Agency 3 (donation forms, procedures), Agency 4 (religious law and rules), Agency 5 (case forms, procedures), Agency 6 (application forms, programs offered, procedures) and Agency 7 (business license, business forms, procedures).

Service Providers domain offers information about the agency and the agency's functions. It refers to the public sector agency's profile such as agency name, address, website and core business of the agency. In addition to that, this category also consists of the contact information of the business unit or person in charge in the agency for particular services or function. As such, from the interview result in Table 2, master data that fall under this category are - Agency 1, 2, 3, 6 (corporate information), Agency 4 (religious law experts) and Agency 5 (judges and lawyer profiles).

Prioritization of master data domains in local government agencies

The priority level of master data domains typically associated with risk mitigation, revenue growth, cost reduction, or organisation efficiency (Allen & Delton Cervo, 2015). The study captures the priority level of each master data entity which is rank 1-5 (lower to higher) priority from the interviewees during the focus group interview. Table 3 describes the master data entities and the priority levels that were recorded from the focus group interviews.

Table 3. Master Data entities and the priority level in local government agencies

Agency	Id	Master Data entities	Master Data Domains	Priority Level (Rank 1-5)
Agency 1	MD1	Corporate Information	Services and products	5
	MD2	Citizens' Profile in the state	Customers	5
	MD3	Procedures	Services and products	5
Agency 2	MD4	Corporate Information	Service providers	4
	MD5	Payment Forms	Services and products	5
	MD6	Payers' Profile	Customers	5
	MD7	Receivers' Profile	Customers	5
	MD8	Procedures	Services and products	5
Agency 3	MD9	Corporate Information	Service providers	3

	MD10	Donation Forms	Services and products	5
	MD11	Donators' Profile	Customers	3
	MD12	Donation Receivers' Profile	Customers	5
	MD13	Asset Developers	Customers	3
	MD14	Procedures	Services and products	5
Agency 4	MD15	Religious Law & Rules	Services and products	5
	MD16	Religious Law experts	Service providers	5
Agency 5	MD17	Judges' Profile	Service providers	5
	MD18	Lawyers' Profile	Service providers	4
	MD19	Plaintiffs' Profile	Customers	5
	MD20	Defendants' Profile	Customers	5
	MD21	Case Forms	Services and products	5
	MD22	Procedures	Services and products	5
Agency 6	MD23	Corporate Information	Service providers	4
	MD24	Application Forms	Services and products	5
	MD25	Programs Offered	Services and products	5
	MD26	Lecturers' Profile	Customers	4
	MD27	Students' Profile	Customers	5
	MD28	Procedures	Services and products	5
	MD29	Students' Feedback/Complaints	Customers	3
Agency 7	MD30	Customers' Profile	Customers	5
	MD31	Contractors' Profile	Customers	4
	MD32	Business License	Services and products	5
	MD33	Business Forms	Services and products	4
	MD34	Businesses' Profile	Customers	4
	MD35	Procedures	Services and products	5
	MD36	Customers' Feedback/Complaints	Customers	4

Overall, based on the focus group interview analyzed result in **Table 3**, master data domains in public sector score high ranks of priority level which is above 3 (three). In particular, the highest priority of the master data domain in the public sector is services and products domain. The mean of priority level for this domain is 4.6. This is truly to the fact and common definition that government roles are to provide services and products across all society (Thenint & Miles, n.d.). Thus, it is important to give a priority in managing the information on how the agencies provide these services and products such as the procedures of the offered services, services access channel and all related forms.

The next important master data domain is the Customers domain. The mean of priority level for this domain is 4.33. The Customers domain is primarily aimed to achieve better cross-functional discipline, quality control, and standardization of an agency's customer identity data and profiling in order to achieve an accurate and shared source of truth about an existing or prospective customer. Finally, the Service Providers domain scores the least mean as compared to other domains. The mean of priority level for this domain is 4.16.

Privacy classification of master data domains in local government agencies

The focus group interviews also gathered the information of privacy classification of each master data entity. Table 4 presents the analyzed results of the focus group interviews in defining master data and their privacy classification.

Table 4. Master Data entities and the privacy classification in local government agencies

Agency	Id	Master Data entities	Privacy Classification
Agency 1	MD1	Corporate Information	Open
	MD2	Citizens' Profile in the state	Confidential
	MD3	Procedures	Open
Agency 2	MD4	Corporate Information	Open
	MD5	Payment Forms	Open
	MD6	Payers' Profile	Confidential
	MD7	Receivers' Profile	Confidential
	MD8	Procedures	Open
Agency 3	MD9	Corporate Information	Open
	MD10	Donation Forms	Open
	MD11	Donators' Profile	Confidential
	MD12	Donation Receivers' Profile	Confidential
	MD13	Asset Developers	Confidential
	MD14	Procedures	Open
Agency 4	MD15	Religious Law & Rules	Open
	MD16	Religious Law experts	Open

Agency 5	MD17	Judges' Profile	Confidential
	MD18	Lawyers' Profile	Open
	MD19	Plaintiffs' Profile	Confidential
	MD20	Defendants' Profile	Confidential
	MD21	Case Forms	Open
	MD22	Procedures	Open
Agency 6	MD23	Corporate Information	Open
	MD24	Application Forms	Open
	MD25	Programs Offered	Open
	MD26	Lecturers' Profile	Confidential
	MD27	Students' Profile	Confidential
	MD28	Procedures	Open
	MD29	Students' Feedback/Complaints	Confidential
Agency 7	MD30	Customers' Profile	Confidential
	MD31	Contractors' Profile	Confidential
	MD32	Business License	Open
	MD33	Business Forms	Open
	MD34	Businesses' Profile	Confidential
	MD35	Procedures	Open
	MD36	Customers' Feedback/Complaints	Confidential

Overall, the interview result in Table 4 shows that 20 over 36 (56%) data sets are classified as open and only 16 over 36 (44%) data sets are classified as confidential by the interviewees. In particular, all 15 data sets in Customers domain are classified as confidential, all 14 data sets in Services and Products domain are classified as open, and 6 over 7 (85%) data sets in Service Providers domain are classified as open.

The analysis result is tally to the fact of common definition that government agencies' roles are to provide services and products across all society (Thenint & Miles, n.d.). The information regarding services and products, and service providers are expected to be published publicly to facilitate citizen accessing the information to know further on the offered services and products such as the procedures of the services, services access channel and all related forms.

4. Conclusion

In conclusion, by conducting focus group interviews from seven (7) selected case studies from Malaysia's local government agencies this study determines master data in government agencies as core information of the agency, shareable between agencies and focusing on the information and service delivery to the citizens and generally grouped into three domains; 1) services and products, 2) customers, and 3) service providers. The highest priority of the master data domain in the public sector is Services and Products and followed by the Customers and Service Providers domains. The result of this study is in contrast with most of the current literatures that stated the MDM typically prioritize on the customer data domain as compared to other domains (Cervo & Allen, 2011). This study also indicates that the government has a high potential to share and integrate the open classification of master data as enterprise master data in MDM and publish them to the citizen with worry-free on the privacy issues since more than 50% of the identified master data entities are classified as open data.

The study of prioritization and privacy classification of master data in government agencies has no claim to be complete. Further research is required to extend the definition of master data domains in government agencies (e.g. by conducting more case studies from federal). However, at this present stage, this paper can serve Malaysia's Government and other practitioners utilizing the existing open data as a foundation to establish the MDM and to improve information delivery to the citizen. From an academic perspective, this paper represents a theory for defining prioritization and privacy classification of master data in government agencies within the area of MDM.

Acknowledgments

The study is financially supported by Research University Grant (Vot No. Q.K130000.2538.09H70), Universiti Teknologi Malaysia and Ministry of Education Malaysia.

References

[1] Allen, M., & Delton Cervo. (2015). Multi Domain Master Data Management. Retrieved from http://www.microsoft.com/en-us/sqlserver/solutions-technologies.aspx?navV3Index=0#tab3

[2] Bonnet, P. (2013). Enterprise Data Governance. Enterprise Data Governance: Reference & Master Data Management, Semantic Modeling. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781118622513

[3] Cervo, D., & Allen, M. (2011). Master Data Management in Practice: Achieving True Customer MDM. New Jersey: John Wiley & Sons.

[4] Clifton, C., Kantarcioğlu, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A., & Suciu, D. (2004). Privacy-preserving data integration and sharing. Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery - DMKD '04, 19. http://doi.org/10.1145/1008694.1008698

[5] Creswell, J. W. (2014). Research design: qualitative, quantitative and mixed methods approaches.

[6] Dreibelbis, A. (2008). Enterprise Master Data Management: An SOA Approach to Managing Core Information. Retrieved from http://dl.acm.org/citation.cfm?id=1391248

[7] Gartner. (2003). SAP's MDM Shows Potential, but Is Rated "Caution."

[8] Linstone, H. a, & Turoff, M. (2002). The Delphi Method - Techniques and applications. The Delphi Method - Techniques and Applications, 1–616. http://doi.org/10.2307/1268751

[9] Loshin, D. (2009). Master Data Management. Master Data Management, 43–65. http://doi.org/10.1016/B978-0-12-374225-4.00003-5

[10] Merriam-Webster Dictionary. (2011). Retrieved February 24, 2016, from http://www.merriam-webster.com/dictionary/government

[11] Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. Information Systems and E-Business Management, 10(3), 395–425. http://doi.org/10.1007/s10257-011-0178-0

[12] Ranjit Kumar. (2011). Research Methodology - a step-by-step guide for beginners (3rd Editio). SAGE.Services, P. (2008). End-to-End Efficiency : Public Services, (July 2004).

[13] Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011). Managing one master data – challenges and preconditions. Industrial Management & Data Systems, 111(1), 146–162. http://doi.org/10.1108/02635571111099776

[14] Smith, H. A., & McKeen, J. D. (2008). Developments in Practice XXX : Master Data Management : Salvation Or Snake Oil ? Master Data Management : Salvation Or Snake Oil ? Communications of the Association for Information Systems, 23(1), 4.

[15] Spruit, M., & Pietzka, K. (2014). MD3M: The master data management maturity model. Computers in Human Behavior. http://doi.org/10.1016/j.chb.2014.09.030

[16] Thenint, H., & Miles, I. (n.d.). Innovation Policy Challenges for the 21st Century. (D. Cox & J. Rigby, Eds.). Routledge. Retrieved from http://www.routledge.com/books/details/9780415896122/

[17] W3C(e-Gov). (2009). eGovernment at W3C: improving access to government through better use of the web. Retrieved from http://www.w3.org/2007/eGov